(Bayesian) regression models

Jens Roeser & Mark Torrance

jens.roeser@ntu.ac.uk
mark.torrance@ntu.ac.uk

Nottingham Trent University

Contents

1. Linear regression models (LM)
2. Linear mixed effects models (LMM)
3. Bayesian linear mixed effects models (BLMM)

Simulated data

- Normal distributed data for two conditions *a* and *b*
  - *simDataContinuous*.*R*: Continuous predictor
  - *simData*.*R*: Discrete predictor
  - *simDataSubjRE*: Discrete predictor and by-subject variance added
- Population parameters are known!

→ The underlying effect: β=50

- I.e. difference between condition *a* and *b*
- Lets try to uncover this effect …
- Open file *AntwerpWS*2017.*Rproj*

Linear regression models

- Single level regression model; ordinary least squares (OLS)
- Linear change in the data given a predictor variable
- Predictor can be continuous (e.g. frequency: 0 – 100) or discrete (e.g. frequency: high vs. low)
- Allows multiple predictors

Linear regression models

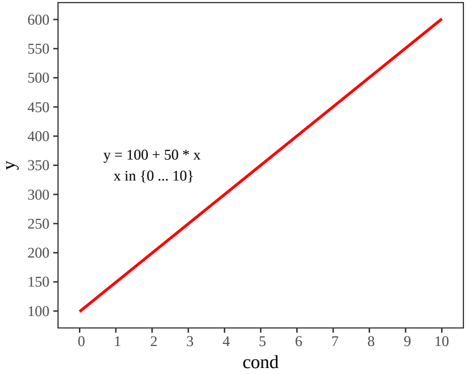
$$y_i = \alpha + \beta \times x_i + \epsilon_i \tag{1}$$

- y: data
- α: intercept
- β: slope (gradient)
- x: predictor
- ϵ: residual error (i.e. noise)

$$\epsilon_i \sim N(0, \sigma^2) \tag{2}$$

- *R* function *lm*() (part of the base *R* package)
- Syntax: *lm*(`outcome variable` ~ `predictor` , `data frame`)
- Your turn: see *R* script *exercise_lm.R*

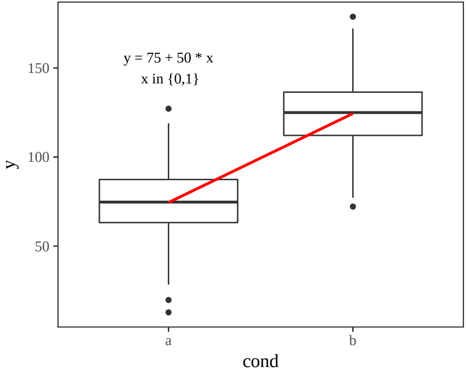
Linear regression models



Predictor x is continuous

```
> head(data)
  cond        y
     9 465.8291
     0 108.5086
     9 569.3380
     7 431.7700
     7 449.9238
     3 233.3673
```

Linear regression models



Predictor *x* is discrete

```
> head(data)
  cond         y
     b 105.10129
     a  82.73471
     b 104.00330
     b  97.21434
     b 101.53576
     a  68.83296
```

Linear regression models

See script *LM_model_discrete.R*

```
> data %>%
+   group_by(cond) %>%
+   summarise (mean = mean(y),
+              sd = sd(y))
# A tibble: 2 x 3
    cond      mean       sd
  <fctr>     <dbl>    <dbl>
1      a  74.49868 18.09908
2      b 124.44140 17.30632
```

Population means are 75 for *cond a* and 125 for *cond b*.

Linear regression models

```
> m <- lm(y ~ cond, data)
> summary(m)$coef
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	74.499	0.804	92.626	< 0.001
condb	49.943	1.120	44.605	< 0.001

- (Intercept): y-value for *cond* = 0; here condition *a*
- `condb`: change from condition *a* (i.e. intercept) to *b*
- condition *b* + intercept is $t \times \mathit{Std.\ Error}$ away from intercept
- Is that what we want?

Linear regression models

- Treatment contrast (default): change from intercept

```
> contrasts(data$cond)
  b
a 0
b 1
```

- Sum contrast (effect magnitude): difference between *a* and *b*

```
> contrasts(data$cond) <- c(-.5, .5)
> colnames(contrasts(data$cond)) <- c("b-a")
> contrasts(data$cond)
   b-a
a -0.5
b  0.5
```

Linear regression models

Table: Treatment contrast

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	74.499	0.804	92.626	< 0.001
condb	49.943	1.120	44.605	< 0.001

Table: Sum contrast

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	99.470	0.560	177.678	< 0.001
condb-a	49.943	1.120	44.605	< 0.001

Linear regression models

- Estimation of effect magnitude
- Linear function can account for and predict unobserved data
- Can account for additional sources of variance and potentially confounding variable by adding those to the model (co-variates)
- However, data are more complex than that:

  E.g. multiple observation per participant/item
- Solution: Linear mixed effects models

Linear mixed effects models

- Extension of linear regression
- Take into account within and between groups variance
- "Mixed": Fixed + random factors
- Fixed: systematic effect
- Random: non-systematic sources of variance; e.g. some participants are faster than others
- *lmer*(); part of *lme*4 (Bates, Mächler, Bolker, & Walker, 2015)

Linear mixed effects models

- New simulated data frame.
- Uncover known parameter $\beta = 50$
- Added by-subjects variance:
  By-subjects intercepts and slopes.
- Models *LM_observations_in_subj.R*
- What do you observe?
- What's the evidence that this parameter is different from 0 (i.e. null hypothesis)?

```
> head(data)
subj  cond         y
   1     a  71.05932
   1     a  62.23969
   1     a  63.93085
   1     a  57.07717
   1     a  56.54501
   1     a  70.29821
```

Linear mixed effects models

```
> m1 <- lm(y ~ cond, data)
> m2 <- lm(ysubjmeans ~ cond, data.subj)
> m3a <- lmer(y ~ cond + (1|subj), data)
> m3b <- lmer(y ~ cond + (1+cond|subj), data)
```

Table: Estimates (see *models_random_effects.R*)

	Estimate	Std. Error	t value	Pr(>|t|)
`m1`	51.189	1.169	43.784	< 0.001
`m2`	51.189	2.823	18.135	< 0.001
`m3a`	51.189	0.860	59.506	
`m3b`	51.189	1.506	33.989	

Linear mixed effects models

```
> m1 <- lm(y ~ cond, data)
> m2 <- lm(ysubjmeans ~ cond, data.subj)
> m3a <- lmer(y ~ cond + (1|subj), data)
> m3b <- lmer(y ~ cond + (1+cond|subj), data)
```

Table: Estimates (see *models_random_effects.R*)

	Estimate	Std. Error	t value	Pr(>|t|)
m1	51.189	1.169	43.784	< 0.001
m2	51.189	2.823	18.135	< 0.001
m3a	51.189	0.860	59.506	
m3b	51.189	1.506	33.989	

Linear mixed effects models

```
> m1 <- lm(y ~ cond, data)
> m2 <- lm(ysubjmeans ~ cond, data.subj)
> m3a <- lmer(y ~ cond + (1|subj), data)
> m3b <- lmer(y ~ cond + (1+cond|subj), data)
```

Table: Estimates (see *models_random_effects.R*)

	Estimate	Std. Error	t value	Pr(>|t|)
`m1`	51.189	1.169	43.784	< 0.001
`m2`	51.189	2.823	18.135	< 0.001
`m3a`	51.189	0.860	59.506	
`m3b`	51.189	1.506	33.989	

Linear mixed effects models

```
> m1 <- lm(y ~ cond, data)
> m2 <- lm(ysubjmeans ~ cond, data.subj)
> m3a <- lmer(y ~ cond + (1|subj), data)
> m3b <- lmer(y ~ cond + (1+cond|subj), data)
```

Table: Estimates (see *models_random_effects.R*)

	Estimate	Std. Error	t value	Pr(>|t|)
m1	51.189	1.169	43.784	< 0.001
m2	51.189	2.823	18.135	< 0.001
m3a	51.189	0.860	59.506	
m3b	51.189	1.506	33.989	

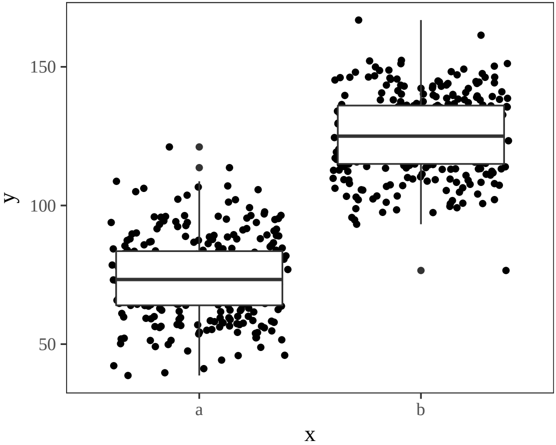
Linear mixed effects models

```
> m1 <- lm(y ~ cond, data)
> m2 <- lm(ysubjmeans ~ cond, data.subj)
> m3a <- lmer(y ~ cond + (1|subj), data)
> m3b <- lmer(y ~ cond + (1+cond|subj), data)
```

Table: Estimates (see *models_random_effects.R*)

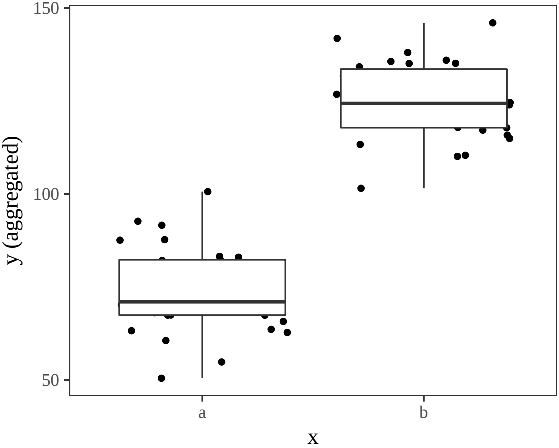
	Estimate	Std. Error	t value	Pr(>|t|)
`m1`	51.189	1.169	43.784	< 0.001
`m2`	51.189	2.823	18.135	< 0.001
`m3a`	51.189	0.860	59.506	?
`m3b`	51.189	1.506	33.989	?

Linear mixed effects models



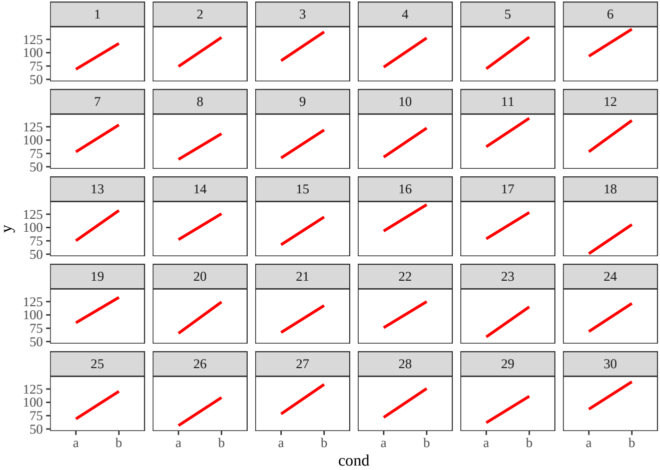
Simulated data

Linear mixed effects models



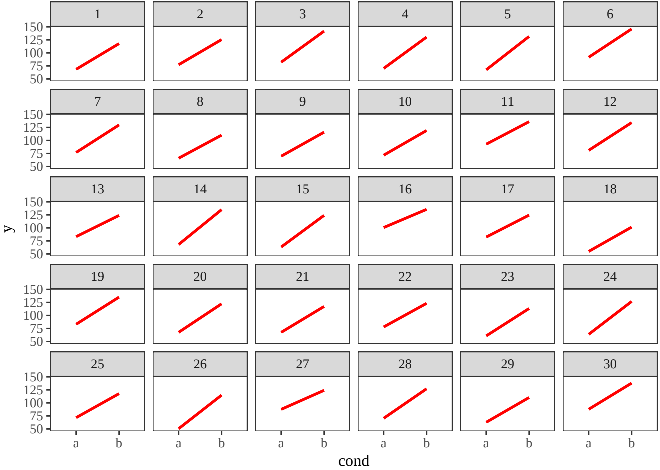
Simulated data, aggregated by subject and condition

Linear mixed effects models



Varying intercepts per subject: $(1|subj)$

Linear mixed effects models



Varying intercepts and slopes per subject: $(1 + cond|subj)$

Linear mixed effects models

- What about p?
- Unclear how to determine df (see Baayen, 2008)
- Alternatives:
  - t-value = 2 as lower bound for $p < 0.05$
  - Satterthwaite approximation *lmerTest* package
  - Model comparison using log likelihood ratio *anova*()
- See *LMER_pvalues.R*

Linear mixed effects models

- *p*-values do **not** tell us whether the difference is large enough to reject the null.
- This relies on the variance within each condition.
- We may reject the null for estimates that are too small to be sensible (e.g. $\sim 5ms$) if the variance is small enough
- …or fail to reject the null for sensible estimates that if the variance is too large.
- It's not about a single value but a range.
- Is 0 a possible value?

Linear mixed effects models

- 95% confidence intervals (CI):

  If we were to repeat our experiment an infinite number of times and calculate a confidence interval each time, 95% of these intervals would contain the true parameter value.
- See simulation: `http://rpsychologist.com/d3/CI/`
- In other words, the estimate is merely the centre of a range of a imaginary range of other intervals which contains the population parameter.
- 5% of these unobserved ranges do not contain the true value.
- See exercises script *LMER95%CIs.R*

Linear mixed effects models

Advantages:

- Flexible models that account for the complexity of data
- Nested data: children nested in classes nested in schools
- Random effects: subject speed varies; effect varies across individuals; slopes and intercepts are correlated
- For a short but thorough intro see Vasishth and Nicenboim (2016)

Linear mixed effects models

Problem:

- Maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013):
  $(1 + cond|subject) + (1 + cond|item)$
- Random effects can be added:

i. $(1 + cond|subject)$ accounts for varying conditional differences across subjects; not plausible in a between subjects design

ii. $(1 + cond|items)$ accounts for varying conditional differences across items; effect is stronger in some items

   Not plausible when there were no matched items

Linear mixed effects models

Problem:

- Maximal random effects structure (Barr et al., 2013):
  $(1 + cond|subject) + (1 + cond|item)$
- Convergence failure: over-parametrisation (Bates, Kliegl, Vasishth, & Baayen, 2015):

  I.e. model is too complex for the data.
- Solution (i): remove random slopes until model converges
- Solution (ii): Bayesian Linear Mixed Effects Models

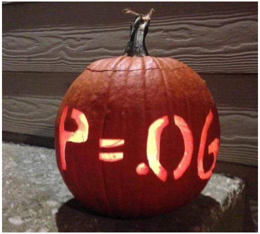
Bayesian linear mixed effects models

- Easier model fit: complex models converge *by definition*
- Answer the question we care about: what's the support for the hypothesis given the data?
- Intuitive interpretation: frequentist estimates are often interpret in a Bayesian manner (Nicenboim & Vasishth, 2016)
- Support for the hypothesis is not quantified by the implausibility of the null.

Bayesian linear mixed effects models

- Easier model fit: complex models converge *by definition*
- Answer the question we care about: what's the support for the hypothesis given the data?
- Intuitive interpretation: frequentist estimates are often interpret in a Bayesian manner (Nicenboim & Vasishth, 2016)
- Support for the hypothesis is not quantified by the implausibility of the null.

Is p the probability that the null is true?
What are we doing with $p = 0.06$?



Bayesian linear mixed effects models

- Bayesian inference based on the **Posterior** distribution approximated from the product of the **Likelihood** and the **Prior**:
  (a) Plausible values for model parameters – **Prior**.
  (b) Probability model of the data generating process – **Likelihood**.
- Sophisticated sampling techniques: Monte Carlo Markov Chain
- Sampling is used to approximate the posterior distribution by creating probability distributions of plausible parameter values.

Bayesian linear mixed effects models

Table: Interpretation of evidence

	NHST*	Bayes
Support for H_1	$P(data \mid H_0)$	$P(H_1 \mid data)$
Inference true effect	95% CI	CrI; HPDI

*Null Hypothesis Significance Testing

Bayesian linear mixed effects models

Table: Interpretation of evidence

	NHST*	Bayes
Support for H_1	$P(data \mid H_0)$	$P(H_1 \mid data)$
Inference true effect	95% CI	CrI; HPDI

- Evidence in favour of H_1:
  - NHST: indirect inference about H_1 based on the (im)plausibility of the data if H_0 were true.
  - Bayes: direct support for H_1 given the data.

*Null Hypothesis Significance Testing

Bayesian linear mixed effects models

Table: Interpretation of evidence

	NHST*	Bayes
Support for H_1	$P(data\|H_0)$	$P(H_1\|data)$
Inference true effect	95% CI	CrI; HPDI

- Intervals containing the true parameter value:
  - NHST: if we were to replicated a experiment a large number of times and calculate a CI each time, 95% of these intervals would include the true parameter value
  - Bayes: probability distribution of possible values for true parameter (e.g. 95% range)

*Null Hypothesis Significance Testing

Bayesian linear mixed effects models

- Probabilistic sampling using Stan – Hamiltonian Monte Carlo
- *R*-Stan interface (Stan Development Team, 2015)
- *R* packages for Bayesian LMMs: *rstanarm* (Gabry & Goodrich, 2016); *brms* (Bürkner, 2017); *rethinking* (McElreath, 2016)

Bayesian linear mixed effects models

See script *BLMM.R*: run the model now

```
m <- stan_lmer(y ~ cond + (1 + cond | subj)
, prior_intercept = student_t(df = 1, location = 0)
, prior = student_t(df = 1, location = 0)
, data = data
, chains = 3
, iter = 1000
, cores = 4
, seed = 17)
```

Bayesian linear mixed effects models

- *prior_intercept*
- *prior_slope*
- *chains*
- *iter*
- *cores*
- *seed*

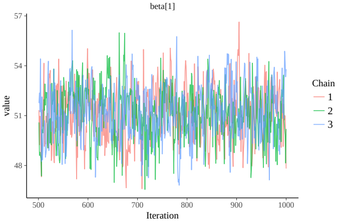
- *student_t* distributions have a *location* parameter and *df*: see script *student-t-distribution*.*R*
- Weakly informative priors: `student_t(df=1)`
- Other priors: *normal*(), *cauchy*(); also on other parameters (e.g. variance-covariance matrix)
- At least 3 chains to determine convergence.
- If model doesn't converge, increase iterations.

Bayesian linear mixed effects models

- Ensuring convergence;
  i.e. model has successfully determined a posterior.
- Compare data to posterior predictive values.
- Traceplots; hairy caterpillars
- $\hat{R} = 1$; Rubin-Gelman statistic (Gelman & Rubin, 1992)
- Example and exercises: *BLMM_modelchecks*.*R*

Bayesian linear mixed effects models

- Ensuring convergence;
  i.e. model has successfully determined a posterior.
- Compare data to posterior predictive values.
- Traceplots; hairy caterpillars
- $\hat{R} = 1$; Rubin-Gelman statistic (Gelman & Rubin, 1992)
- Example and exercises: *BLMM_modelchecks.R*

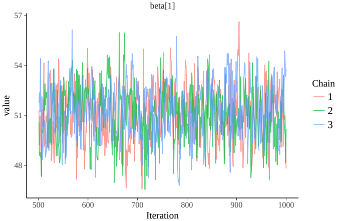


Real

Traceplots

Bayesian linear mixed effects models

- Ensuring convergence;
  i.e. model has successfully determined a posterior.
- Compare data to posterior predictive values.
- Traceplots; hairy caterpillars
- $\hat{R} = 1$; Rubin-Gelman statistic (Gelman & Rubin, 1992)
- Example and exercises: *BLMM_modelchecks.R*

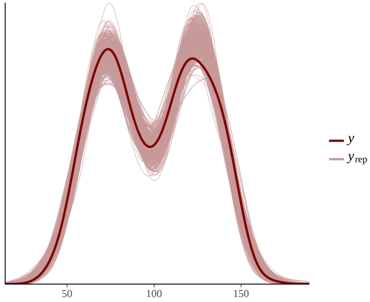


Real

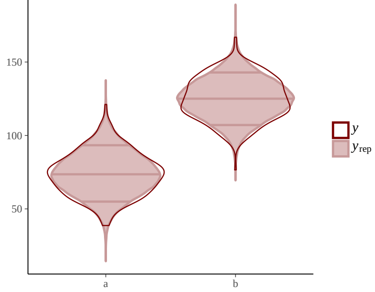
Fake

Traceplots

Bayesian linear mixed effects models



Distribution



By-cond

Comparison of data and posterior predictive values

Bayesian linear mixed effects models

- Does the interval contain 0? Is 0 a possible parameter value?
- 95% credible interval (CrI): range of possible parameter values with equal probability mass assigned to each tail (percentile intervals)
- 95% highest posterior density interval (HPDI): interval that embraces the assigned probability mass; identical to CrI for symmetrically distributed posteriors
- Go through script *BLMM_CrI.R*

Bayesian linear mixed effects models

- Does the interval contain 0? Is 0 a possible parameter value?
- What's the probability that this isn't probable?
- $P(\hat{\beta} < 0)$: proportion of posterior samples that is smaller than 0 $\rightarrow$ probability that parameter is smaller than 0 (i.e. speed-up)
- Go through script *BLMM_postprob*.R

Bayesian linear mixed effects models

- Does the interval contain 0? Is 0 a possible parameter value?
- What's probability of a slow-down/speed-up?
- $P(\hat{\beta} < 0)$: proportion of posterior samples that is smaller than 0
  $\rightarrow$ probability that parameter is smaller than 0 (i.e. speed-up)
- Go through script *BLMM_postprob.R*

Bayesian linear mixed effects models

- What's the most likely value for unknown parameter?
- Maximum A posteriori (MAP): most frequent ~ probable value
- Go through script *BLMM_MAP.R*

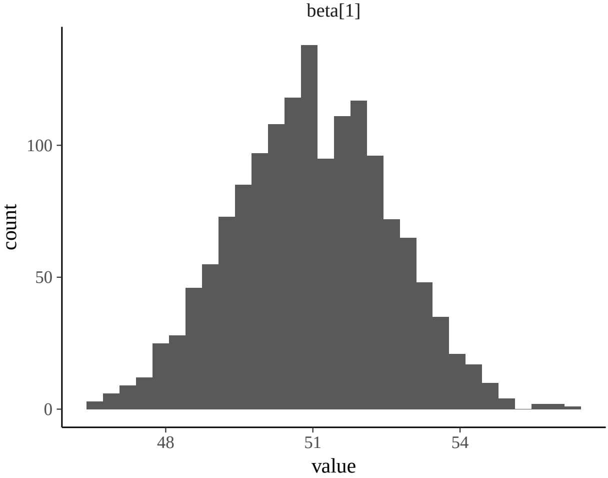
Bayesian linear mixed effects models

- Comparing hypotheses (i.e. models) using Bayes Factors (BF)
- BF_{10}: support for H_1 over H_0
- $BF_{10} = 2$: H_1 is two times more likely than H_0; convincing?
- $BF_{10} = 3$–5: weak to moderate evidence
- $BF_{10} > 10$: strong support
- $BF_{10} < .3$: evidence against H_1
- For evidence in favour of H_0: BF_{01}
- Savage-Dickey density ratio (Dickey, Lientz, et al., 1970)
- see e.g. Baguley (2012), Dienes (2014), Lee and Wagenmakers (2014), Wagenmakers, Lodewyckx, Kuriyal, and Grasman (2010)
- Go through script *BLMM_BF.R*

Bayesian linear mixed effects models

- All you need!
- Model summary: *BLMM_modelsummary.R*

Bayesian linear mixed effects models



Posterior probability distribution of effect $\hat{\beta}$

Bayesian linear mixed effects models

Table: Estimates of BLMM. Evidence strongly supports H_1 ($BF_{10} > 5 \times 10^{34}$)

$\hat{\beta}$	2.5%	97.5%	P($\hat{\beta} < 0$)
51.12	47.98	54.26	< 0.001

Bayesian linear mixed effects models

- The future is Bayes!
- Existing probabilistic sampling software (Jags, Stan, WinBugs, pyMCMC) makes approximation of posterior easily possible.
- Open source access via *R* and *Python*.
- Complex models will converge (easy model fit).
- Answers the questions we ask
- …including support in favour of the null!
- No dichotomisation of the significance of the evidence.
- Probability distributions of possible parameter values.
- Interpretation of evidence is intuitive.
- Custom made models: mixture models, ex-Gaussian

Bayesian linear mixed effects models

- Introductions to using Bayesian linear mixed models:
  - Nicenboim and Vasishth (2016): applying *rstanarm* to psycholinguistic data
  - Sorensen, Hohenstein, and Vasishth (2016): building LMMs in Stan
- Bayesian theory:
  - Great books; with *R* code: Kruschke (2014), McElreath (2016)
  - Very technical; focus on hierarchical models: Gelman et al. (2014)

References I

Baayen, R. H. (2008). *Analyzing linguistic data. a practical introduction to statistics using r*. Cambridge: Cambridge University Press.

Baguley, T. (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*. Basingstoke: Palgrave Macmillan.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

Dickey, J. M., Lientz, B. P., et al. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a markov chain. *The Annals of Mathematical Statistics*, *41*(1), 214–226.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(781), 1–17.

Gabry, J., & Goodrich, B. (2016). rstanarm: Bayesian applied regression modeling via stan [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=rstanarm` (R package version 2.13.1)

References II

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457–472.

Kruschke, J. K. (2014). *Doing bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

McElreath, R. (2016). *Statistical rethinking: A bayesian course with examples in R and Stan*. CRC Press.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas – Part II. *arXiv preprint arXiv:1602.00245*.

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists. *Quantitative Methods for Psychology*, *12*(3), 175–200.

Stan Development Team. (2015). *Stan modeling language user's guide and reference manual.* `http://mc-stan.org/`. (R package version 2.6.1.)

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas – Part I. *arXiv preprint arXiv:1601.01126*.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive psychology*, *60*(3), 158–189.